

Evaluation of Surrogate Endpoints in HIV Clinical Trials

Michael D. Hughes

Harvard School of Public Health

Talk Outline: Experiences in 2 Meta-analyses

- Focus on surrogacy in the context of intermediate endpoints for evaluating treatment in clinical trials
- Rationale for investigating surrogacy across clinical trials
- Conceptual approach
- Examples: evaluation of HIV RNA and CD4 as surrogate endpoints for AIDS/death
- Some issues

Rationale for Evaluating Surrogacy Across Trials

- Limitations of “PTE” (and other within-trial methods?):
 - Poor precision unless highly significant treatment differences on clinical outcome
 - Trials which show no such difference (even if well-powered) don't contribute information about surrogacy
 - Measurement error hampers evaluation of surrogacy at individual patient level
 - Conceptual issues - e.g. “proportion” not bounded by 0 and 1

Rationale for Evaluating Surrogacy Across Trials

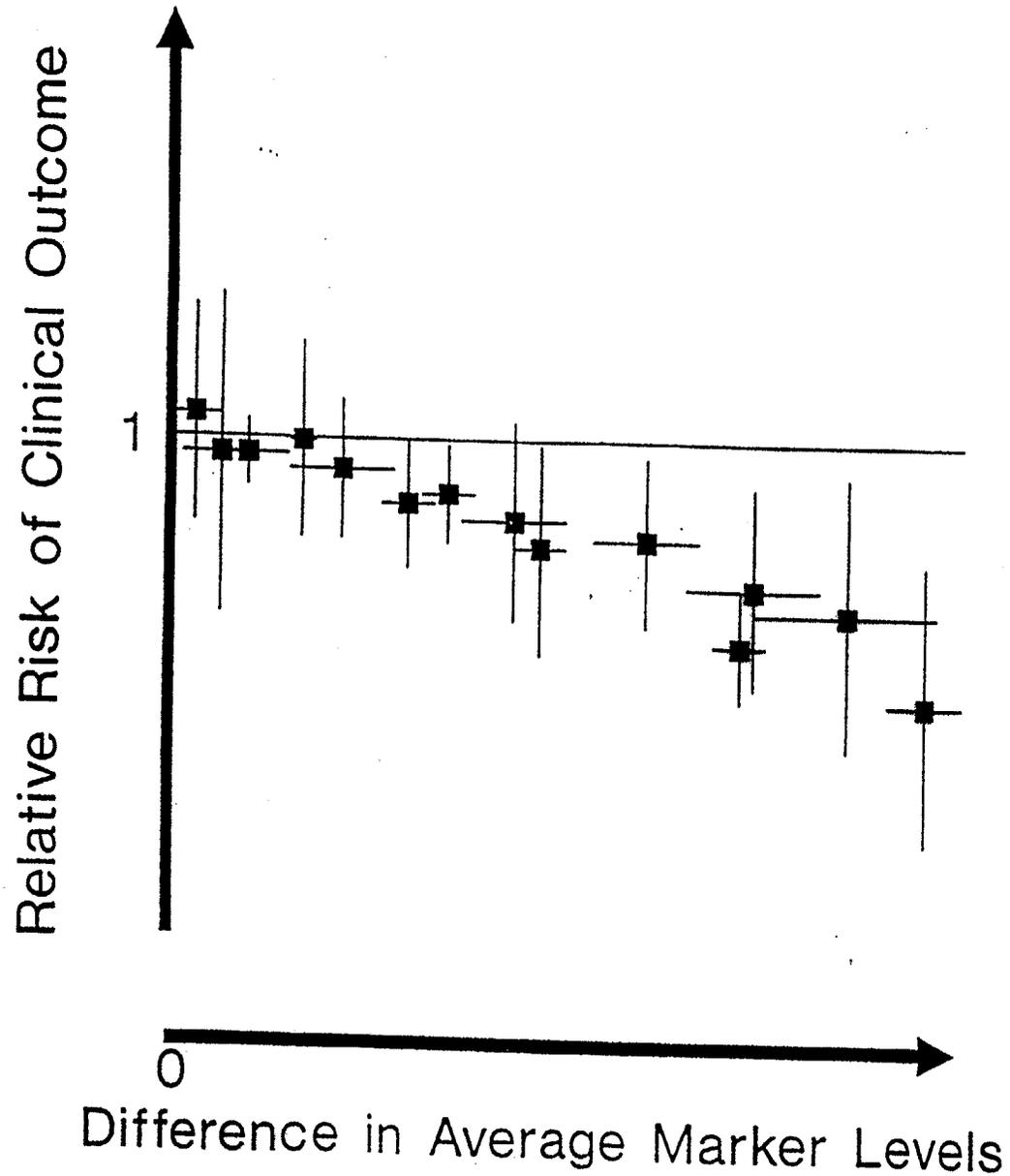
- Advantages of cross-study approaches:
 - Consistency across studies (and treatments) is critical - want to avoid reporting bias problem
 - Want to answer the question: Given an observed difference on a response variable, what is the likely difference in clinical outcome - focus on “intermediate” rather than “concurrent” surrogacy
 - Reduce impact of model selection & measurement error; use information from studies showing no difference in clinical outcome

Conceptual Approach

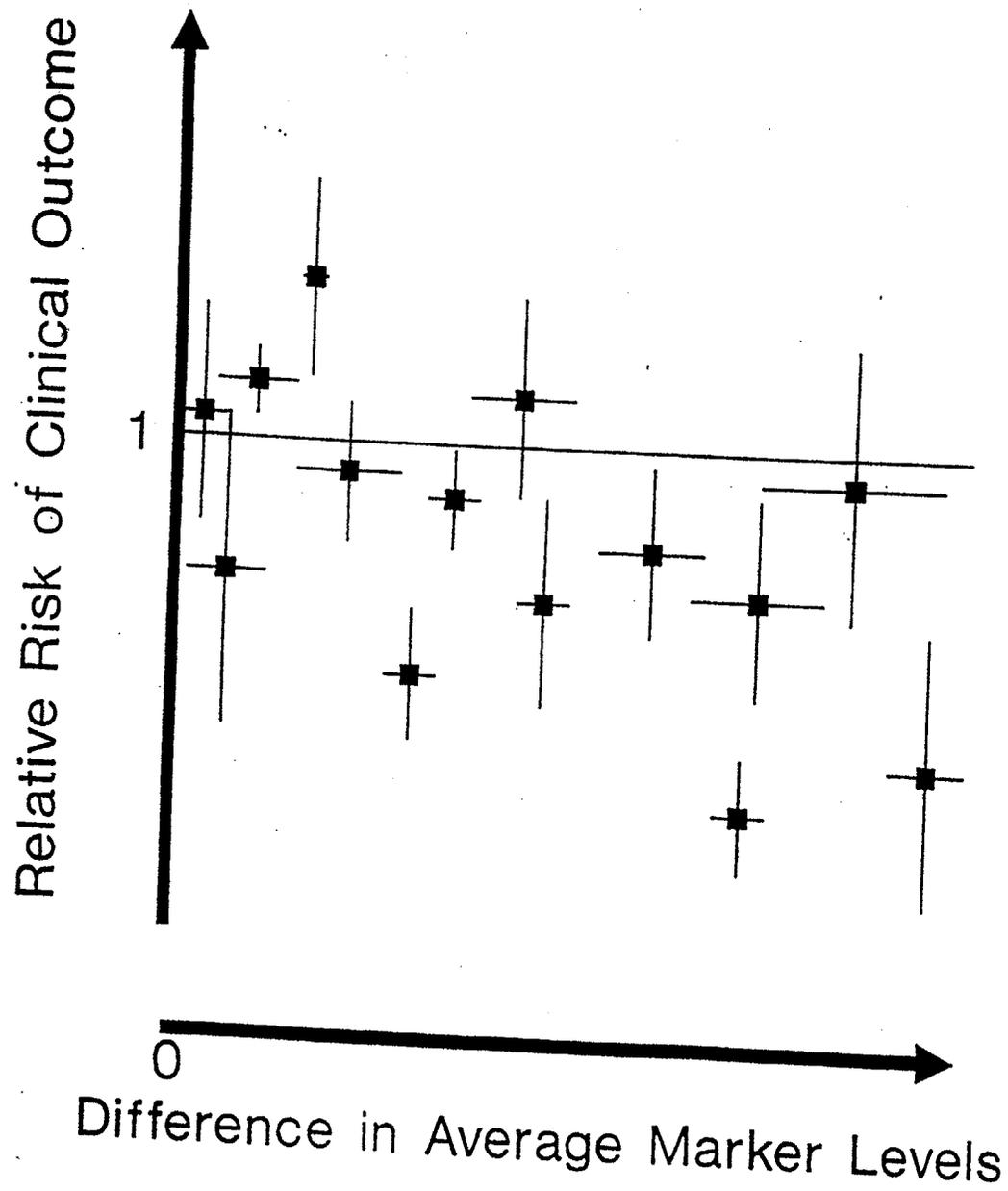
- Evaluate the association between differences between treatments in their effect on a response variable and the corresponding differences in effect on clinical outcome
 - across clinical trials

Ref: A'Hern (1988) Br. J. Cancer
Hughes et al. (1995) J. AIDS

Evaluating Surrogate Endpoints: Meta-analysis A Good Response Variable for Prediction



Evaluating Surrogate Endpoints: Meta-analysis A Poor Response Variable for Prediction



Evaluating Surrogate Endpoints: Meta-analysis Statistical Model

- In i th clinical trial:

$\gamma_i =$ true difference between treatments for response variable

$\theta_i =$ for clinical outcome

- Observe $\hat{\gamma}_i$ and $\hat{\theta}_i$

- Model (two levels):

Within-trial: describes variability of $(\hat{\gamma}_i, \hat{\theta}_i)$ about true (γ_i, θ_i)

Between-trial: describes association between γ_i and θ_i

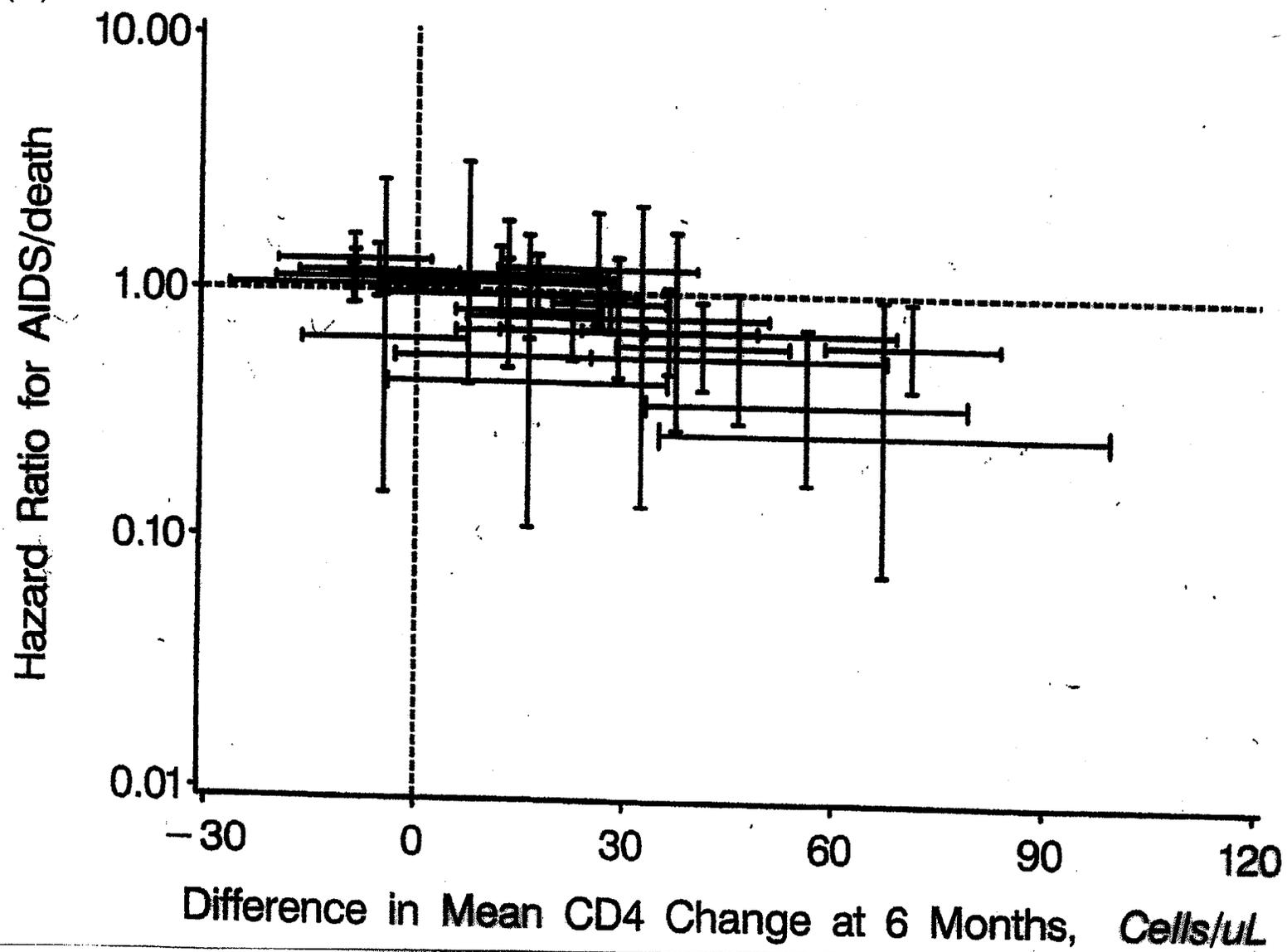
Evaluating Surrogate Endpoints: Meta-analysis Simple Statistical Model

- **Between-trial:** $\theta_i = \alpha + \beta\gamma_i + \varepsilon_i$ where $\varepsilon_i \sim N(0, \tau^2)$
- $\beta=0$ ---> **Response variable has no predictive value**
- $\beta \neq 0, \tau^2=0$ ---> **Treatment effect on clinical outcome can be predicted perfectly given true effect on response variable**
- **Good rationale for having $\alpha=0$**
 - **since if $\alpha \neq 0$, then having true difference on response variable of zero does not imply no difference in clinical outcome.**
[i.e. $\alpha \neq 0$ violates Prentice criteria]

Example (1): Evaluation of HIV RNA & CD4 as intermediate endpoints

- All trials of nRTIs (international collaboration of pharmaceutical companies and trials organizations)
 - with HIV RNA measurements
 - completed by Sept 30, 1997
 - >6 months duration
 - ≥ 1 AIDS/death event
- Are treatment differences in AIDS/death over 2 years of follow-up associated with treatment differences in mean change in HIV RNA or CD4 cell count at 24 weeks?

(a)



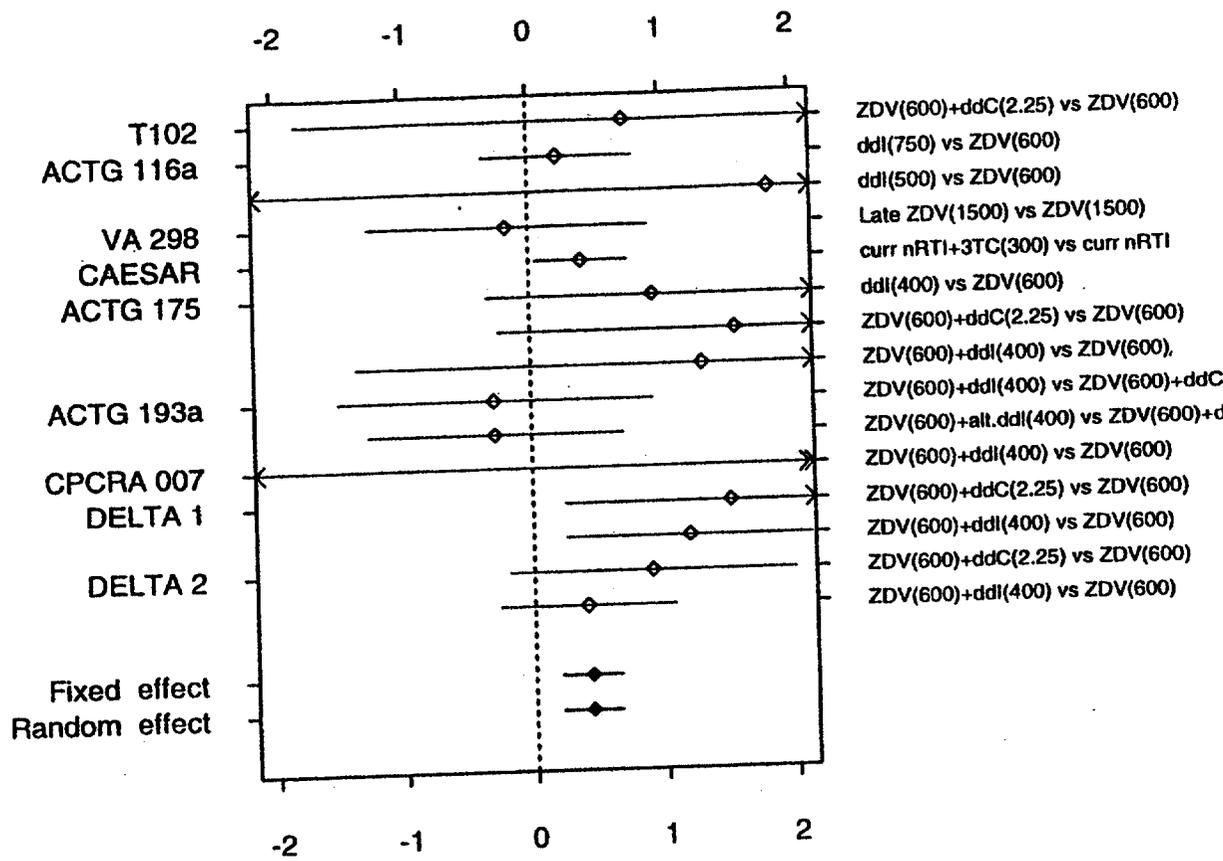
Example (2): Evaluation of CD4 as an intermediate endpoint

- All trials of antiretroviral therapies undertaken by the Adult AIDS Clinical Trials Group
 - completed by Jan 1, 1997
 - >6 months duration
 - ≥ 1 AIDS/death event
- Are treatment differences in AIDS/death (or death alone) over 2 years of follow-up associated with treatment differences in mean change in CD4 cell count at 6 months?

[Ref: Daniels and Hughes (1997) Stat Med; Hughes et al (1998) AIDS]

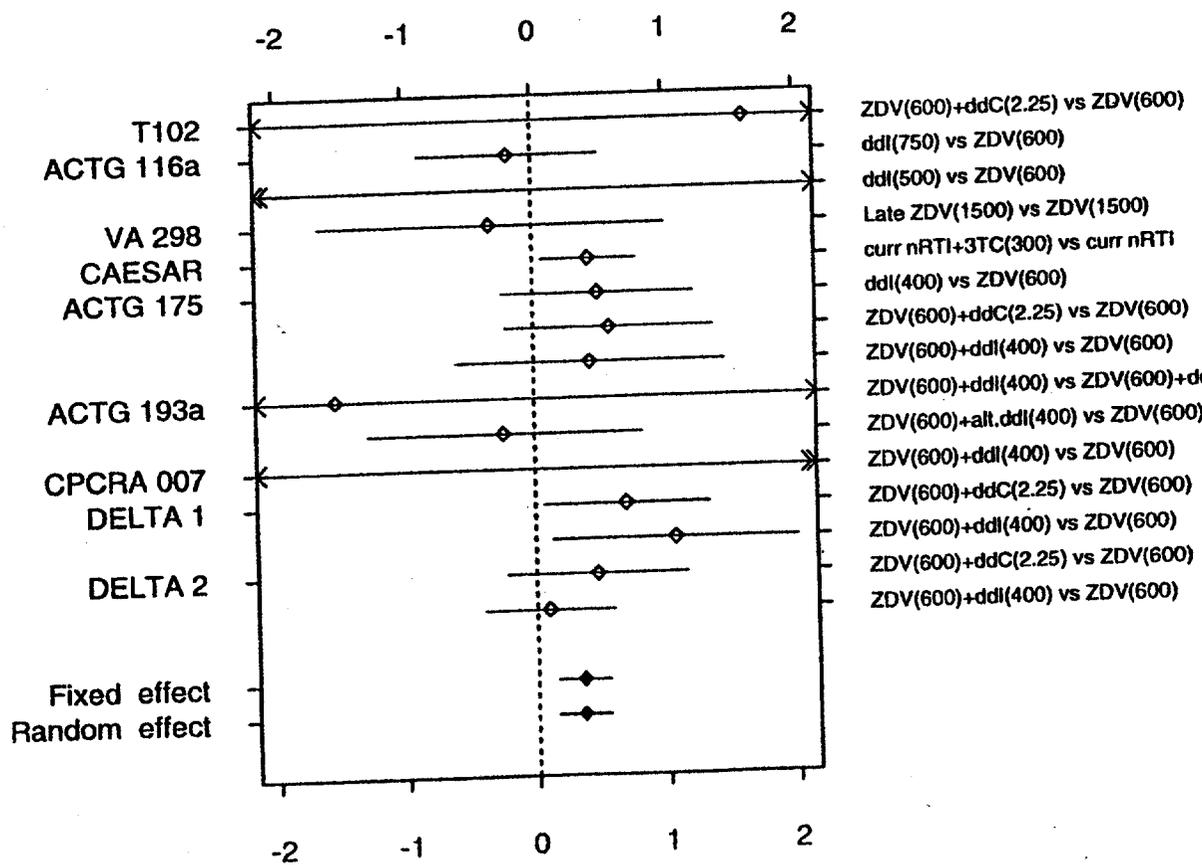
<u>Markers Included</u>		PTE (95% CI)
Week 24	AUCMB	
RNA	-	0.43 (0.21, 0.66)
-	RNA	0.53 (0.26, 0.81)
CD4	-	0.35 (0.15, 0.55)
-	CD4	0.24 (0.07, 0.41)
RNA & CD4	-	0.55 (0.28, 0.83)
-	RNA & CD4	0.63 (0.31, 0.95)

AUCMB: Area under the curve minus baseline; CI: confidence interval.



Proportion of treatment effect explained

(b)

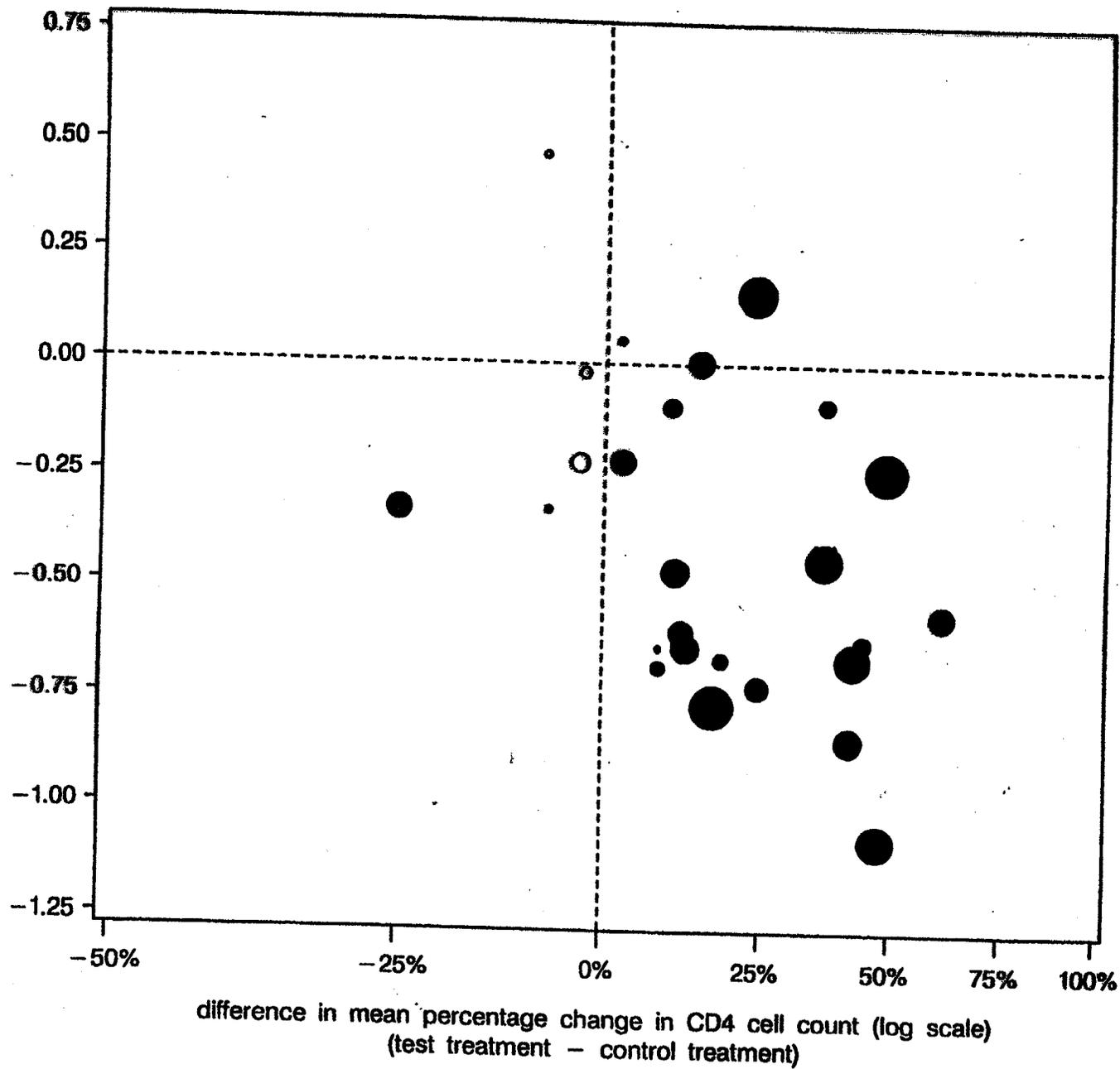


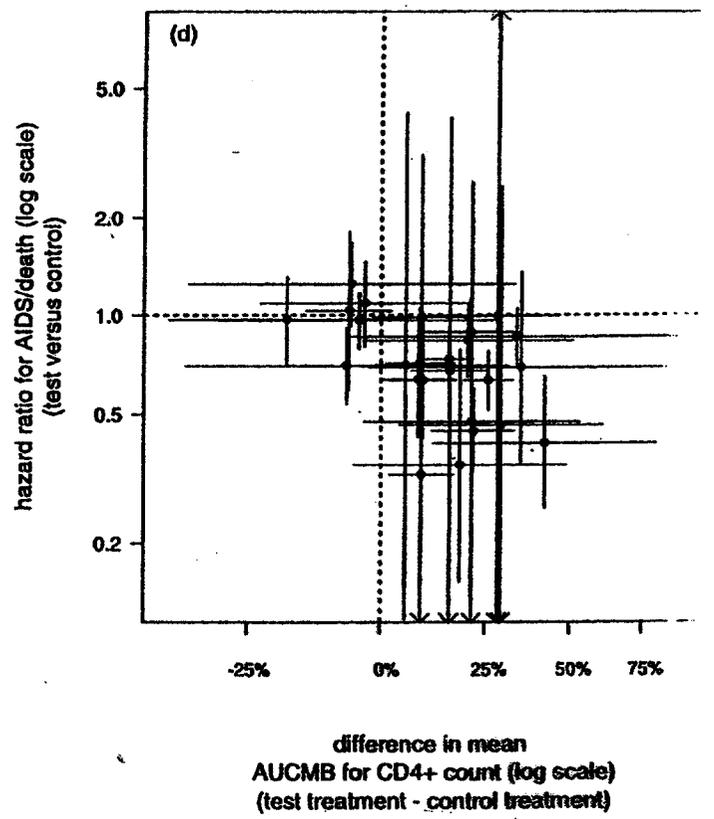
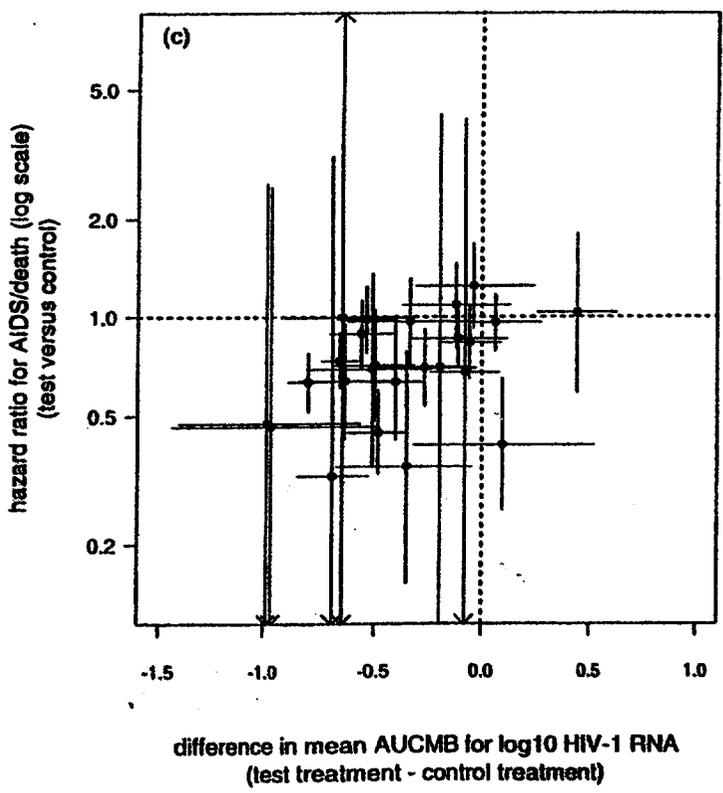
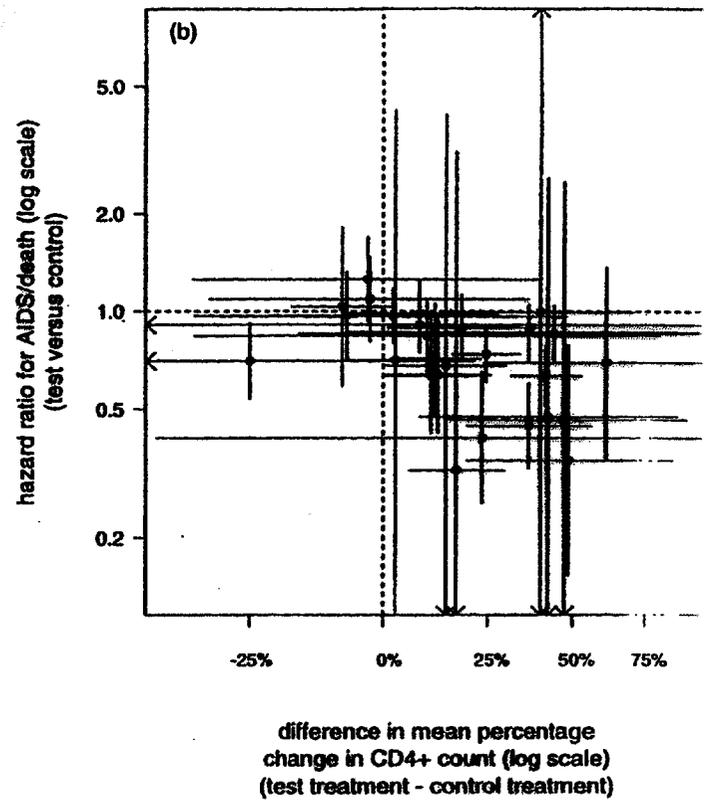
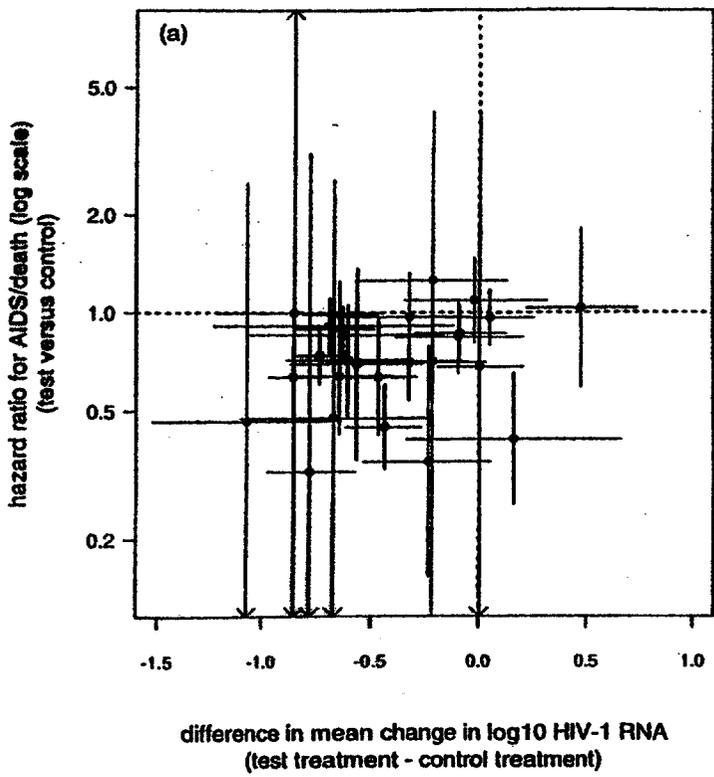
An aside PTE

HIV RNA & CD4 Meta-analysis

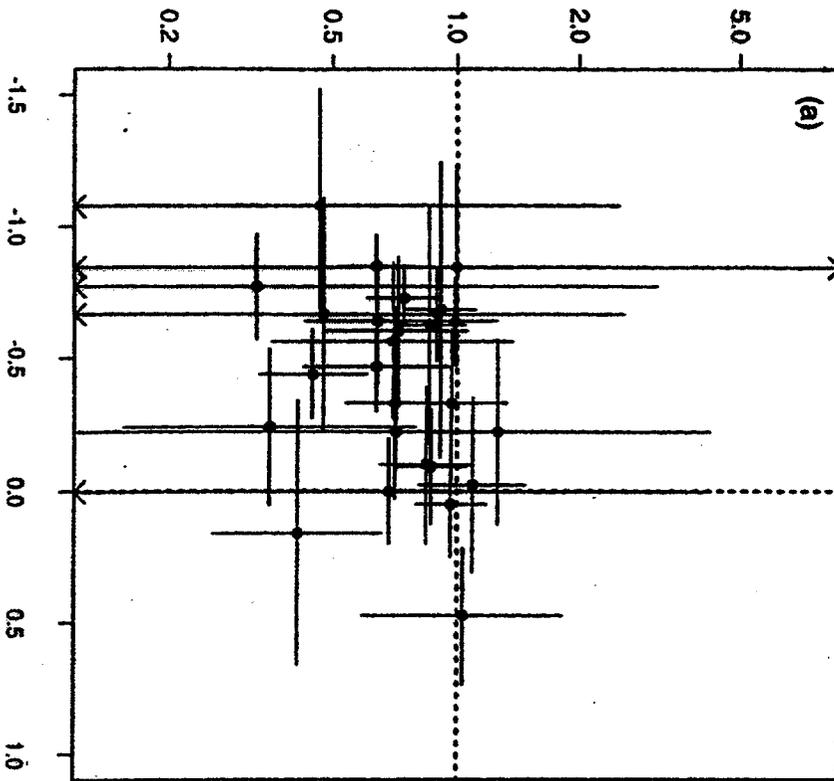
- Reasonable concordance between differences in effect on HIV RNA and CD4, and difference in effect on AIDS/death for this class of drugs.
- CD4 may be better as an intermediate endpoint than HIV RNA – but no clear winner in multivariate analysis
 - Trials should use both
- Difference between treatments of about 50+ CD4 cells/ μ l necessary before reasonably certain of a corresponding clinical benefit
- Similarly, need difference in AUCMB of 1+ log HIV RNA

difference in mean change in HIV-1 log₁₀ RNA
(test treatment - control treatment)



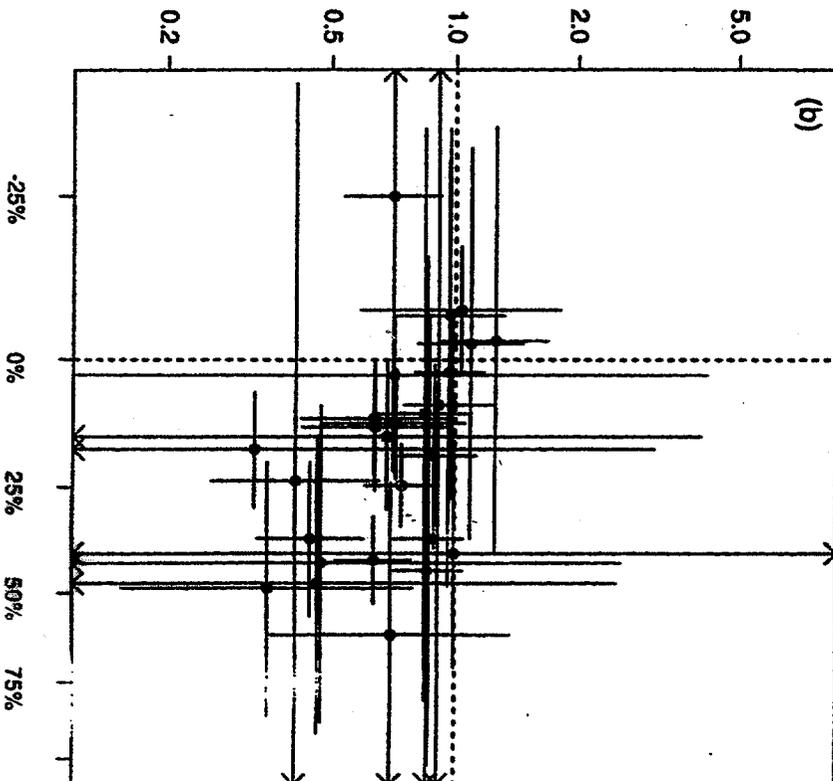


hazard ratio for AIDS/death (log scale)
(test versus control)



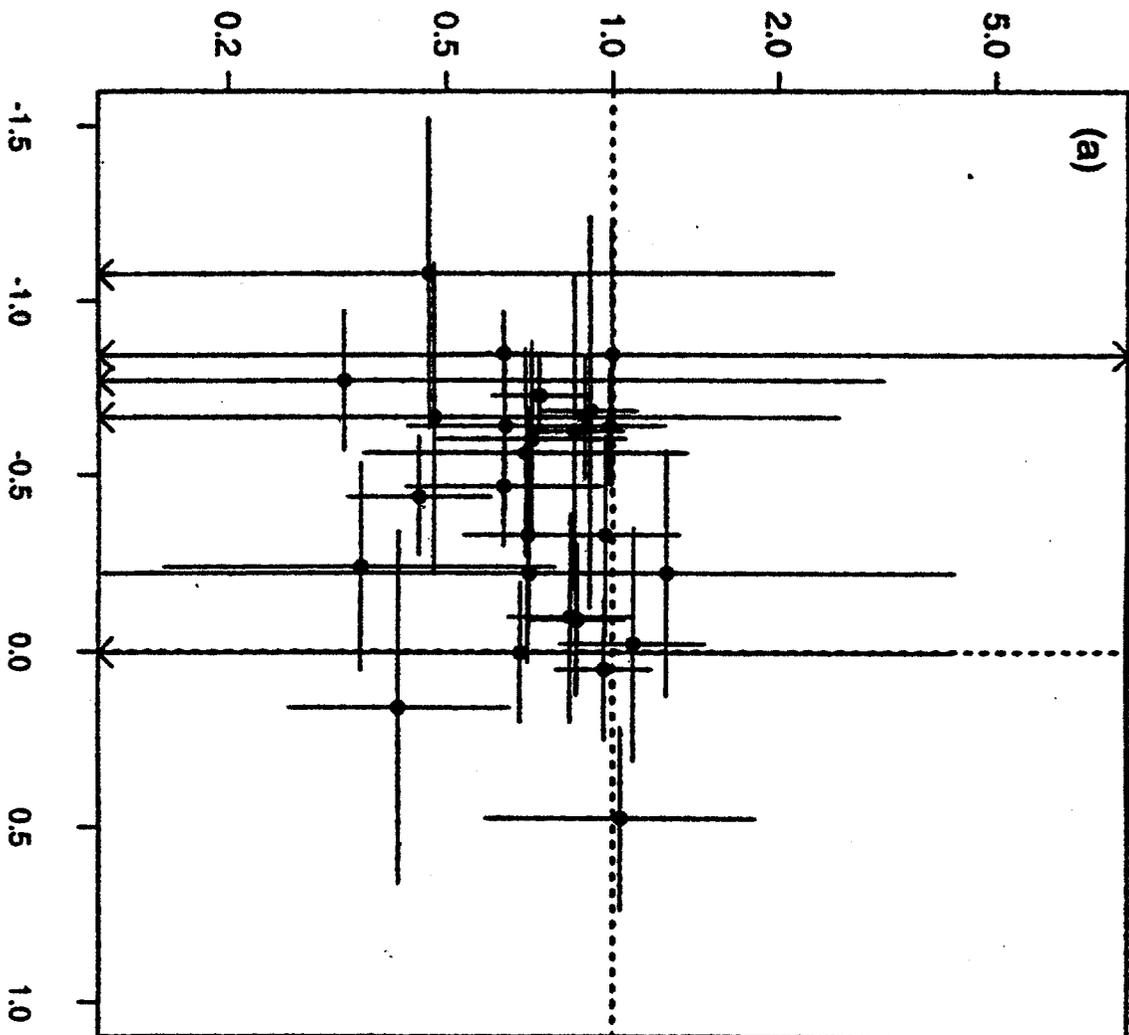
difference in mean change in log₁₀ HIV-1 RNA
(test treatment - control treatment)

hazard ratio for AIDS/death (log scale)
(test versus control)



difference in mean percentage
change in CD4+ count (log scale)
(test treatment - control treatment)

hazard ratio for AIDS/death (log scale)
(test versus control)



difference in mean change in log₁₀ HIV-1 RNA
(test treatment - control treatment)

Model fitting: Priors

- $\alpha \sim N(0, A_\alpha)$
 $\beta \sim N(0, A_\beta)$
 $\gamma_i \sim N(0, A_{\gamma_i})$ } A 's large \rightarrow "noninformative"

- τ^2 trickier

- I (DuMouchel): $\pi(\tau^2) = \frac{\sigma_c}{(\sigma_c + \tau)^2} \cdot \frac{1}{2\tau}$

- II (shrinkage): $\pi(\tau^2) = \frac{\sigma_c^2}{\sigma_c^2 + \tau^2}$

- III (flat): $\pi(\tau^2) = \text{const.}$

where σ_c^2 is harmonic mean of σ_i^2 's

Handling 3-armed trials

Within-trial level becomes.....

$$\begin{bmatrix} \hat{\theta}_{i1} \\ \hat{\theta}_{i2} \\ \hat{\delta}_{i1} \\ \hat{\delta}_{i2} \end{bmatrix} \sim N \left(\begin{bmatrix} \theta_{i1} \\ \theta_{i2} \\ \delta_{i1} \\ \delta_{i2} \end{bmatrix}, \begin{pmatrix} \sigma_{i1}^2 & \rho_{i0}\sigma_{i1}\sigma_{i2} & \rho_{i1}\sigma_{i1}\delta_{i1} & \rho_{i2}\sigma_{i1}\delta_{i2} \\ & \sigma_{i2}^2 & \rho_{i2}\sigma_{i2}\delta_{i1} & \rho_{i1}\sigma_{i2}\delta_{i2} \\ & & \delta_{i1}^2 & \rho_{i0}\delta_{i1}\delta_{i2} \\ & & & \delta_{i2}^2 \end{pmatrix} \right)$$

- ρ_{i0}, ρ_{i1} from Cox models/linear reg.
- Now have $\rho_{i1} + \rho_{i2}$

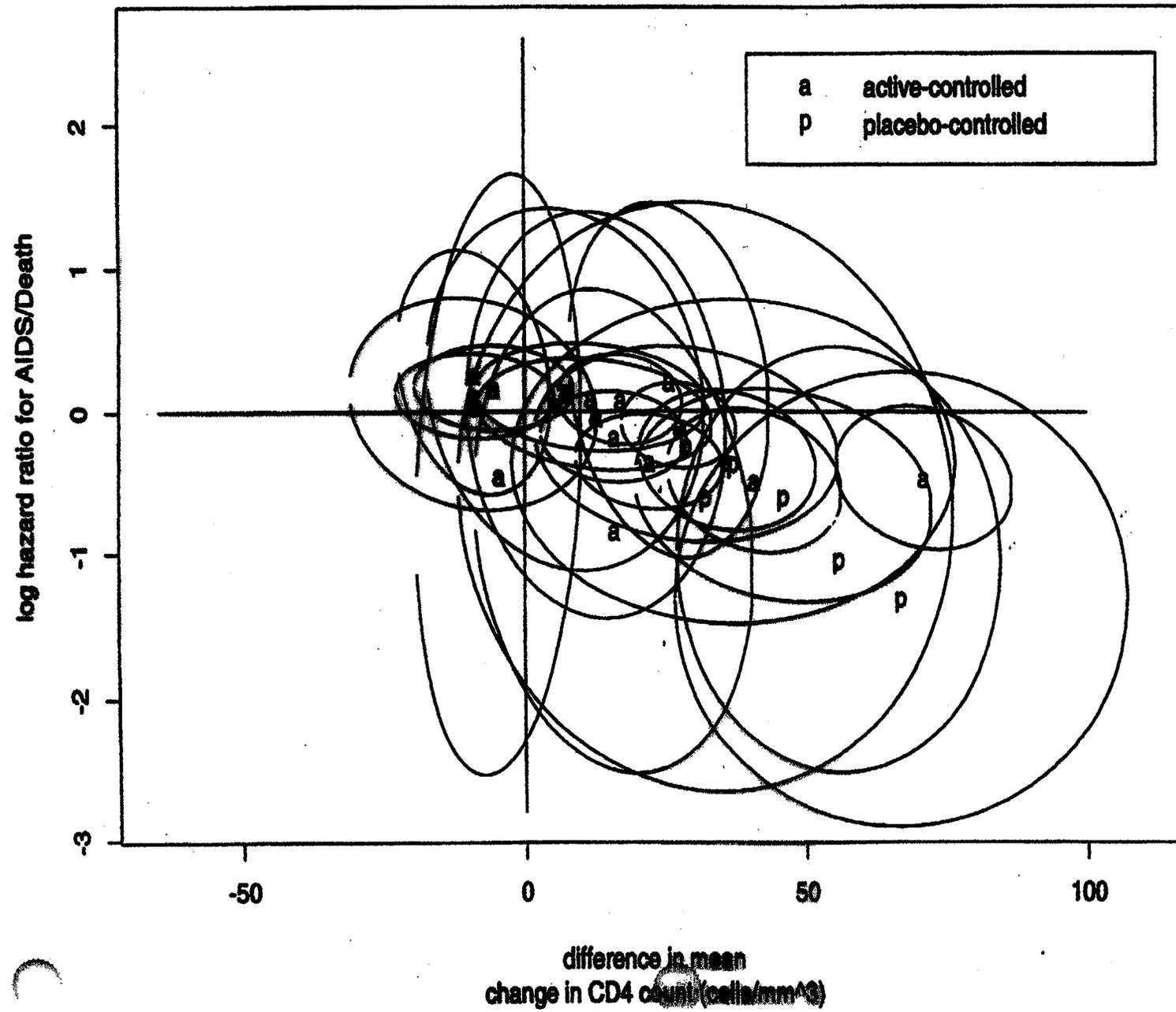
Table I. Treatment differences for the log hazard ratio for the development of AIDS or death over 2 years (θ_i) and the difference in mean change in CD4 cell count between baseline and 6 months ($\hat{\gamma}_i$) for studies of the AIDS Clinical Trial Group ($\hat{\theta}_i$, $\hat{\delta}_i$, $\hat{\rho}_i$ are estimates of the standard error of θ_i , the standard error of $\hat{\gamma}_i$, and the correlation between θ_i and $\hat{\gamma}_i$)

Study	Reference	Test treatment*	Standard treatment*	$\hat{\theta}_i(\hat{\delta}_i)$	$\hat{\gamma}_i(\hat{\delta}_i)$	$\hat{\rho}_i$
002	31	ZDV[600]	ZDV[1500]	0.048 (0.092)	-9.2 (9.0)	-0.14
016	32	ZDV[1200]	placebo	-1.035 (0.370)	56.0 (11.8)	-0.02
019a	33	ZDV[1500]	placebo	-0.235 (0.282)	28.8 (11.0)	-0.13
		ZDV[500]	placebo	-0.594 (0.307)	46.1 (10.7)	-0.15
019b	34	ZDV[1500]	placebo	-1.313 (0.651)	67.1 (16.8)	0.01
		ZDV[500]	placebo	-0.359 (0.465)	37.2 (16.3)	-0.00
036	35	ZDV[1500]	placebo	-0.598 (0.707)	32.2 (18.0)	-0.06
112	†	ddC[†]	ZDV[†]	-0.447 (0.732)	-4.7 (6.1)	0.17
114	†	ddC[2.25]	ZDV[600]	0.267 (0.121)	-9.1 (5.6)	-0.22
116a	36	ddI[750]	ZDV[600]	0.096 (0.156)	11.8 (8.4)	-0.15
		ddI[500]	ZDV[600]	-0.022 (0.161)	12.8 (8.6)	-0.19
116b	37	ddI[750]	ZDV[600]	0.180 (0.130)	15.9 (5.3)	-0.07
		ddI[500]	ZDV[600]	-0.355 (0.137)	22.2 (5.4)	-0.11
118	38	ddI[200]	ddI[750]	0.112 (0.121)	-8.9 (5.8)	-0.06
		ddI[500]	ddI[750]	0.166 (0.120)	-5.5 (5.8)	-0.05
119	39	ddC[2.25]	ZDV[600]	-0.035 (0.340)	12.8 (9.5)	-0.08
155	40	ZDV/ddC[600/2.25]	ZDV[600]	-0.102 (0.121)	27.5 (4.2)	-0.09
		ddC[2.25]	ZDV[600]	0.083 (0.129)	17.1 (4.5)	-0.10
175	41	ZDV/ddC[600/2.25]	ZDV[600]	-0.348 (0.202)	36.1 (6.5)	-0.13
		ZDV/ddI[600/400]	ZDV[600]	-0.467 (0.207)	71.2 (6.4)	-0.17
		ddI[400]	ZDV[600]	-0.487 (0.207)	40.9 (6.4)	-0.19
229	42	ZDV/SQV[600/1800]	ZDV/ddC[600/2.25]	0.148 (0.518)	7.3 (10.2)	-0.13
		ZDV/ddC/SQV[600/2.25/1800]	ZDV/ddC[600/2.25]	-0.841 (0.680)	15.9 (10.2)	-0.16
241	43	ZDV/ddI/NVP[600/400/400]	ZDV/ddI[600/400]	0.211 (0.258)	25.8 (7.3)	-0.17

* Figures in brackets are the total daily dose (mg), ZDV = zidovudine, ddI = didanosine, ddC = zalcitabine, NVP = nevirapine, SQV = saquinavir

† Not published

‡ The ddC dose was weight dependent and the ZDV dose depended on the dose of ZDV being received prior to study entry



Model fitting

- Combining over within- + between-trial levels of model:

$$\begin{pmatrix} \hat{\theta}_i \\ \hat{\delta}_i \end{pmatrix} \sim N \left[\begin{pmatrix} \alpha + \beta \gamma_i \\ \delta_i \end{pmatrix}, \begin{pmatrix} \sigma_i^2 + \tau^2 & \rho_i \sigma_i \delta_i \\ \rho_i \sigma_i \delta_i & \delta_i^2 \end{pmatrix} \right]$$

- fit using "Bayesian" methods
- data required: $\hat{\theta}_i, \hat{\delta}_i$
- approximation (c.f. DuMouchel):
use $\hat{\sigma}_i^2 + \hat{\delta}_i^2$ instead of $\sigma_i^2 + \delta_i^2$
- complexity: ρ_i
→ with patient data, bootstrap to obtain $\hat{\rho}_i$ + use instead of ρ_i .

Model

- N trials (can involve different treatment comparisons)

- i th trial:

δ_i = true difference between treatments w.r.t. response variable

θ_i = w.r.t. true endpoint

Assume:

$$\begin{pmatrix} \hat{\theta}_i \\ \hat{\delta}_i \end{pmatrix} \sim N \left[\begin{pmatrix} \theta_i \\ \delta_i \end{pmatrix}, \begin{pmatrix} \sigma_i^2 & \rho_i \sigma_i \delta_i \\ \rho_i \sigma_i \delta_i & \delta_i^2 \end{pmatrix} \right]$$

$$\theta_i | \delta_i \sim N[\alpha + \beta \delta_i, \tau^2]$$

δ_i fixed effects

Model fitting

- Gibbs Sampler (+ Metropolis algorithm)

- Bayes factors used to assess $H_0: \beta = 0$
[+ $H_0: \alpha = 0$].

- use Savage Dickey density ratio
with a proper prior for β similar to
the unit information prior of Kass +
Wasserman.

→ a normal prior, mean 0

+ variance = number of t/t comparisons
x covariance matrix of
parameter estimates
from the WLS regression

- Predictions from $f(\theta_i | \hat{\gamma}_i, \delta_i^2, \alpha, \beta, \tau^2)$
 $= N[(\alpha + \beta \hat{\gamma}_i), \tau^2 + \beta^2 \delta_i^2]$

→ use MCMC output: $f(\theta_i | \hat{\gamma}_i, \delta_i^2) = \sum_{m=1}^M f(\theta_i | \hat{\gamma}_i, \delta_i^2, \tau^2)$

Parameter Estimates : Median [2.5%, 97.5%]
from posterior distⁿ

Model	Parameters	DuMouchel Prior I*	Shrinkage Prior II*	Flat Prior III*
A	α	.072 [-0.030, .181]	.072 [-.038, .189]	.071 [-.045, .194]
	β	-.010 [-.014, -.006]	-.010 [-.015, -.006]	-.010 [-.015, -.005]
	τ^2	.0009 [.0000, .0161]	.0040 [.0002, .0268]	.0063 [.0002, .0432]
B	α	0	0	0
	β	-.008 [-.012, -.005]	-.009 [-.012, -.005]	-.009 [-.013, -.005]
	τ^2	.0012 [.0000, .0172]	.0047 [.0001, .0266]	.0070 [.0002, .0428]

Checking predictive value for individual trials

- Exclude a treatment comparison, j , from meta-analysis
- Fit model without j th comparison
- Obtain prediction for $\hat{\theta}_j$ given the observed $\hat{\delta}_j$, $\hat{\delta}_j^2$, $\hat{\sigma}_j^2$, & prediction int.
- Is $\hat{\theta}_j$ within prediction int.?

If more $\hat{\theta}_j$'s outside their prediction intervals than expected, then may be evidence against surrogacy of response variable for all drugs.

Sensitivity Analysis

Comparison excluded

θ_i for comparison excluded

Trial	Test Treatment	Standard Treatment	β	r^2	θ_i for comparison excluded		Z-score*
					observed ($\hat{\theta}_i$)	predicted median 95% interval	
002	ZDV[600]	ZDV[1500]	-.009	.0054	.05	.08 [-.22,.39]	.19
016	ZDV[1200]	placebo	-.008	.0043	-1.04	-.44 [-1.23,.34]	1.47
019a	ZDV[1500]	placebo	-.008	.0047	-.24	-.24 [-.88,.37]	-.02
	ZDV[500]	placebo	-.008	.0046	-.59	-.37 [-1.05,.29]	.66
019b	ZDV[1500]	placebo	-.008	.0044	-1.31	-.55 [-1.81,.79]	1.15
	ZDV[500]	placebo	-.009	.0045	-.36	-.31 [-1.30,.66]	.09
036	ZDV[1500]	placebo	-.008	.0045	-.60	-.24 [-1.74,1.12]	.47
112	ddC[**]	ZDV[**]	-.009	.0046	-.45	.06 [-1.38,1.49]	.69
114	ddC[2.25]	ZDV[600]	-.008	.0037	.27	.08 [-.22,.38]	-1.23
116a	ddI[750]	ZDV[600]	-.009	.0047	.10	-.11 [-.49,.26]	-1.06
	ddI[500]	ZDV[600]	-.008	.0050	-.02	-.10 [-.50,.28]	-.42
116b	ddI[750]	ZDV[600]	-.009	.0054	-.18	-.14 [-.47,.19]	.24
	ddI[500]	ZDV[600]	-.008	.0043	-.36	-.18 [-.54,.15]	.98
118	ddI[200]	ddI[750]	-.009	.0050	.11	.08 [-.23,.40]	-.20
	ddI[500]	ddI[750]	-.009	.0053	.17	.05 [-.25,.36]	-.73
119	ddC[2.25]	ZDV[600]	-.009	.0046	-.04	-.10 [-.83,.57]	-.21
155	ZDV/ddC[600/2.25]	ZDV[600]	-.009	.0057	-.10	-.24 [-.58,.07]	-.87
	ddC[2.25]	ZDV[600]	-.009	.0048	.08	-.14 [-.48,.19]	-1.37
175	ZDV/ddC[600/2.25]	ZDV[600]	-.009	.0052	-.35	-.30 [-.77,.18]	.20
	ZDV/ddI[600/400]	ZDV[600]	-.010	.0046	-.47	-.70 [-1.28,-.16]	-.86
	ddI[400]	ZDV[600]	-.008	.0045	-.49	-.34 [-.83,.12]	.59
229	ZDV/SQV[600/1800]	ZDV/ddC[600/2.25]	-.009	.0045	.15	-.06 [-1.06,.96]	-.40
	ZDV/ddC/SQV[600/2.25/1800]	ZDV/ddC[600/2.25]	-.008	.0047	-.84	-.16 [-1.50,1.21]	.99
241	ZDV/ddI/NVP[600/400/400]	ZDV/ddI[600/400]	-.009	.0043	.21	-.24 [-.79,.31]	-1.59

* Z-score = $\frac{\hat{\theta}_i - \hat{\theta}_{pred}}{\sqrt{\text{var}(\hat{\theta}_{pred})}}$ where $\hat{\theta}_{pred}$ and $\text{var}(\hat{\theta}_{pred})$ are the predicted mean and variance, respectively.

** the ddC dose was weight dependent and the ZDV dose depended on the dose of ZDV being received prior to study entry.

* using shrinkage prior

Predicting $\theta_i | \hat{\gamma}_i, \delta_i$

- useful in practice
 - selecting drugs for further study
 - "accelerated" approval
- allows assessment of sensitivity to choice of prior for τ^2

Predictions : median [2.5%, 97.5%] $\theta_i / \delta_i, \delta_i$

$\hat{\gamma}_i$	Prior*	$\delta_i = 0$	$\delta_i = 5$	$\delta_i = 10$	$\delta_i = 15$
0	I	.00 (-.12,.12)	.00 (-.14,.14)	.00 (-.21,.21)	.00 (-.29,.29)
	II	.00 (-.18,.18)	.00 (-.19,.19)	.00 (-.24,.24)	.00 (-.31,.31)
	III	.00 (-.22,.22)	.00 (-.23,.23)	.00 (-.28,.28)	.00 (-.34,.34)
10	I	-.09 (-.22,.05)	-.09 (-.24,.07)	-.09 (-.31,.12)	-.09 (-.39,.20)
	II	-.09 (-.27,.09)	-.09 (-.28,.11)	-.09 (-.34,.15)	-.09 (-.41,.23)
	III	-.09 (-.31,.14)	-.09 (-.33,.15)	-.09 (-.37,.18)	-.09 (-.44,.25)
20	I	-.17 (-.31,-.03)	-.17 (-.34,-.01)	-.17 (-.40,.04)	-.17 (-.48,.11)
	II	-.17 (-.36,.01)	-.17 (-.38,.03)	-.17 (-.44,.07)	-.17 (-.51,.14)
	III	-.17 (-.41,.05)	-.17 (-.42,.07)	-.17 (-.47,.10)	-.17 (-.53,.16)
30	I	-.25 (-.42,-.10)	-.25 (-.44,-.08)	-.25 (-.51,-.04)	-.25 (-.58,.03)
	II	-.26 (-.46,-.06)	-.26 (-.48,-.05)	-.26 (-.54,-.01)	-.26 (-.61,.05)
	III	-.26 (-.50,-.02)	-.26 (-.52,-.01)	-.26 (-.57,.02)	-.26 (-.64,.07)
40	I	-.34 (-.52,-.16)	-.34 (-.55,-.15)	-.34 (-.62,-.11)	-.34 (-.69,-.05)
	II	-.34 (-.57,-.13)	-.34 (-.59,-.12)	-.34 (-.64,-.08)	-.34 (-.72,-.03)
	III	-.34 (-.61,-.09)	-.34 (-.63,-.08)	-.34 (-.68,-.05)	-.34 (-.75,-.01)
50	I	-.42 (-.64,-.21)	-.42 (-.67,-.21)	-.42 (-.72,-.18)	-.42 (-.80,-.13)
	II	-.43 (-.68,-.19)	-.43 (-.71,-.18)	-.43 (-.75,-.15)	-.43 (-.82,-.10)
	III	-.43 (-.72,-.15)	-.43 (-.74,-.14)	-.43 (-.78,-.13)	-.43 (-.85,-.08)
60	I	-.51 (-.76,-.27)	-.51 (-.78,-.26)	-.51 (-.84,-.24)	-.51 (-.91,-.20)
	II	-.51 (-.80,-.25)	-.51 (-.82,-.24)	-.51 (-.87,-.22)	-.51 (-.93,-.18)
	III	-.52 (-.84,-.21)	-.52 (-.86,-.21)	-.52 (-.90,-.18)	-.52 (-.96,-.15)

* Prior I: $\pi(\tau^2) = \frac{\sigma_c}{(\sigma_c + \tau)^2} \frac{1}{2\tau}$; Prior II: $\pi(\tau^2) = \frac{\sigma_c^2}{(\sigma_c^2 + \tau^2)^2}$; Prior III: $\pi(\tau^2) = d\tau^2$

from model with $\alpha = 0$

Table 2. Predicted true hazard ratios and 95% prediction intervals for progression to AIDS or death during 2 years for selected observed differences in mean change in CD4 cell count from baseline to 6 months, and the standard error (SE) of the difference, for test versus control treatments.

Observed difference in mean change in CD4 cells ($\times 10^6/l$)	Predicted hazard ratio	95% Prediction interval				
		SE = 0	SE = 5	SE = 10	SE = 15	SE = 20
Progression to AIDS/death						
0	1.07	0.88-1.32	0.87-1.36	0.81-1.45	0.75-1.57	0.69-1.72
10	0.97	0.80-1.18	0.79-1.21	0.74-1.28	0.68-1.39	0.61-1.53
20	0.88	0.72-1.07	0.71-1.09	0.66-1.16	0.60-1.25	0.55-1.36
30	0.80	0.65-0.97	0.63-0.99	0.59-1.05	0.54-1.12	0.49-1.23
40	0.72	0.57-0.90	0.56-0.91	0.52-0.96	0.48-1.02	0.44-1.11
50	0.65	0.50-0.83	0.49-0.84	0.46-0.87	0.43-0.93	0.39-1.01
60	0.59	0.44-0.77	0.43-0.78	0.41-0.81	0.37-0.85	0.34-0.92
Death						
0	1.06	0.69-1.61	0.68-1.63	0.67-1.68	0.65-1.77	0.62-1.88
10	0.97	0.65-1.47	0.64-1.48	0.62-1.52	0.59-1.59	0.56-1.68
20	0.90	0.59-1.36	0.58-1.37	0.57-1.40	0.54-1.46	0.51-1.53
30	0.83	0.53-1.29	0.53-1.29	0.51-1.30	0.48-1.34	0.45-1.41
40	0.77	0.48-1.21	0.47-1.21	0.46-1.24	0.43-1.26	0.40-1.31
50	0.71	0.42-1.17	0.42-1.17	0.40-1.18	0.38-1.19	0.36-1.22
60	0.65	0.37-1.15	0.37-1.15	0.35-1.15	0.33-1.16	0.31-1.16

Issues (1)

- Collaboration required!
- Resources - meta-analyses require extensive data management, programming, statistical support but are often difficult to fund
- Differences in trial design for example:
 - definitions of AIDS events
 - length of follow-up
 - marker measurement schedules,
 - assays used

Issues (2)

- Conceptual framework:
 - include or exclude clinical events prior to time of marker measurement?
 - use difference in mean marker values only or use mean marker value of each treatment?
 - should it matter whether you look at (A-B) treatment comparison or (B-A)?
 - which metrics of measurement should be used?
-- clinical outcome and marker effects.

Issues (3)

- Statistical methodology:
 - complex robustness?
 - model validation
 - make better use of patient-level data?